

DEVICE AND METHOD FOR DETECTING SPEAKER'S HAND POSITION

Publication number: JP2001175869

Publication date: 2001-06-29

Inventor: KYO TOCHU; RYU SEIKETSU; KIN CHIEN; KIN SHOYO; SEO YANG-SEOCK

Applicant: SAMSUNG ELECTRONICS CO LTD

Classification:

- international: G06T1/00; G06K9/00; G06T7/00; G06T7/40; G06T7/60; H04N9/64; G06T1/00; G06K9/00; G06T7/00; G06T7/40; G06T7/60; H04N9/64; (IPC1-7): G06T7/00; G06T7/60; H04N9/64

- European: G06K9/00F1

Application number: JP20000248307 20000818

Priority number(s): KR19990055577 19991207

Also published as:

US6611613 (B1)
KR20010054674 (A)
NL1016006C (C2)

Report a data error here

Abstract of JP2001175869

PROBLEM TO BE SOLVED: To provide a speaker's hand position detecting device and its method which can accurately and speedily search for the positions of the eyes and face of a speaker through video and are tolerant of video noise. **SOLUTION:** This device includes an eye position detecting means 10 which detects a pixel showing a relatively strong gray feature in video and decides as a candidate area an area showing partial cohesion characteristics and having a texture feature in the area that the pixel forms, a face position determining means 20 which determines an optimum search template among search templates formed matching a position extracted from the candidate area of the eye, and an extraction position stabilizing means 30 which determines the position area of the speaker by using the optimum search template.



Data supplied from the esp@cenet database - Worldwide

(19) 日本国特許庁 (J P)

(12) 公開特許公報 (A)

(11) 特許出願公開番号

特開2001-175869

(P2001-175869A)

(43) 公開日 平成13年6月29日 (2001.6.29)

(51) Int.Cl. ⁷	識別記号	F I	データ* (参考)
G 0 6 T 7/00	3 0 0	G 0 6 T 7/00	3 0 0 D 5 B 0 5 7
1/00	3 4 0	1/00	3 4 0 A 5 C 0 6 6
7/60	2 0 0	7/60	2 0 0 Z 5 L 0 9 6
H 0 4 N 9/64		H 0 4 N 9/64	J

審査請求 未請求 請求項の数32 O L (全 16 頁)

(21) 出願番号 特願2000-248307(P2000-248307)

(22) 出願日 平成12年8月18日 (2000.8.18)

(31) 優先権主張番号 99-55577

(32) 優先日 平成11年12月7日 (1999.12.7)

(33) 優先権主張国 韓国 (K R)

(71) 出願人 390019839

三星電子株式会社

大韓民国京畿道水原市八達区梅蔭洞416

(72) 発明者 姜 東 中

大韓民国 京畿道 水原市 八達区 靈通

洞 1053-2番地 豊林2団地アパート

235棟 1903号

(72) 発明者 柳 聖 杰

大韓民国 京畿道 水原市 八達区 梅蔭

4洞 205-22番地

(74) 代理人 100064414

弁理士 磯野 道造

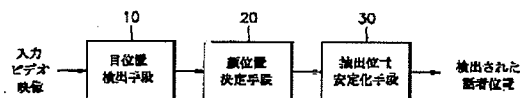
最終頁に続く

(54) 【発明の名称】 話し手位置検出装置及びその方法

(57) 【要約】

【課題】 映像で話し手の目と顔の位置をより正確かつより迅速に探すことができ、しかも映像雑音にも強い話し手位置検出装置及びその方法を提供する。

【解決手段】 映像で、相対的に強い灰色特徴を示すピクセルを検出し、このピクセルが形成する領域中から部分凝集特性を示しかつテクスチャ特徴を有する領域を目の候補領域とする目位置検出手段10と、前記目の候補領域中から抽出した位置にマッチングさせて形成した探索テンプレートから最適の探索テンプレートを決定する顔位置決定手段20と、この最適の探索テンプレートを利用して話し手の位置領域を決定する抽出位置安定化手段30とを含む。



【特許請求の範囲】

【請求項1】 入力されたRGB映像で、相対的に強い灰色の特徴を示すピクセルを検出し、この検出されたピクセルが形成する領域の中から部分凝集特性を示してテクスチャ特徴を有する領域を目的候補領域として決定する目位置検出手段と、

前記目的候補領域の中から抽出した2個の領域の位置にマッチングさせて探索テンプレートを形成し、RGB映像で前記探索テンプレートの領域内におけるピクセルの色度値に対する確率距離の和を正規化した値と、前記探索テンプレートによって推定される目、口及び鼻の位置における水平エッジの大きさをを用いて形成された探索テンプレート中、最適の探索テンプレートを決定する顔位置決定手段と、

前記最適の探索テンプレートを利用して最小外廓四角形を構成し、形状メモリに記憶された各ピクセルに対してカウント値の中から前記最小外廓四角形の領域に対応するカウント値については増加させ、前記最小外廓四角形以外の領域は対応するカウント値については減少させることによって、前記形状メモリで所定の値以上のカウント値が位置した領域を話し手位置領域として出力する抽出位置安定化手段と、を含むことを特徴とする話し手位置検出装置。

【請求項2】 前記目位置検出手段は、入力されたRGB映像を解析して相対的に強い灰色の特徴を示すピクセルを抽出する強灰色抽出部と、前記強灰色抽出部によって抽出されたピクセル中、相互に連結されたピクセル同士を結合して領域を形成する領域形成部と、前記領域形成部によって形成された各領域に対して部分凝集特性を検出する領域形状解析部と、前記領域形成部によって形成された各領域に対してテクスチャ特徴を検出するテクスチャ抽出部と、前記領域形成部によって形成された領域の中から前記部分凝集特性及び前記テクスチャ特徴が各々所定の値より大きい領域を目的候補領域と判別する目候補判別部と、を備えることを特徴とする請求項1に記載の話し手位置検出装置。

【請求項3】 前記顔位置決定手段は、予め備えられたモデルテンプレートを、目的候補領域の中から抽出した2個の領域の位置に対応させて相似変換することによって入力されたRGB映像上に探索テンプレートを形成する顔テンプレート形成部と、前記RGB映像で探索テンプレートの領域内におけるピクセルの色度値に対する確率距離の和を前記探索テンプレートの大きさに対して正規化した正規化確率距離を算出する確率距離演算部と、前記探索テンプレートで推定された目、鼻及び口の位置で入力されたRGB映像の水平エッジ特徴値を検出するエッジ特徴解析部と、

前記顔テンプレート形成部によって形成された複数の探索テンプレート中、前記正規化確率距離及び前記水平エッジ特徴値に所定の重みを与えて得られた加重値によって最適のテンプレートを決定する最適テンプレート決定部と、を備えることを特徴とする請求項1に記載の話し手位置検出装置。

【請求項4】 前記抽出位置安定化手段は、前記入力されたRGB映像の大きさに対応するピクセル数のカウント値を記憶する形状メモリと、前記最適の探索テンプレート内で顔映像が含まれる最小外廓四角形を構成する最小外廓四角形構成部と、前記形状メモリに記憶された各ピクセルに対してカウント値の中から前記最小外廓四角形の領域に対応するカウント値については増加させ、前記最小外廓四角形以外の領域に対応するカウント値については減少させる形状メモリ更新部と、前記形状メモリで所定の値以上のカウント値が位置する領域を話し手位置領域として出力する追跡位置抽出部と、を備えることを特徴とする請求項1に記載の話し手位置検出装置。

【請求項5】 前記強灰色抽出部は、前記RGB映像の各ピクセルに対してカラーを示す色要素の最大値と最小値との差が所定の値より小さく、かつ最大値が他の所定の値より小さく分布する場合に、相対的に強い灰色性向を示すピクセルとして抽出することを特徴とする請求項2に記載の話し手位置検出装置。

【請求項6】 前記領域形状解析部は、各領域の環状性値を算出する環状性解析部を備え、前記目候補判別部は、この各領域の環状性値が所定の値より小さい場合に、その領域を目的候補領域から取り除くことを特徴とする請求項2に記載の話し手位置検出装置。

【請求項7】 前記領域形状解析部は、各領域の細長比を算出する細長比解析部を備え、前記目候補判別部は、この各領域の細長比が所定の値より大きいのか、他の所定の値より小さい場合に、その領域を目的候補領域から取り除くことを特徴とする請求項2に記載の話し手位置検出装置。

【請求項8】 前記領域形状解析部は、各領域に対して全体映像の大きさに相対的な領域の大きさを算出する領域大きさ解析部を備え、前記目候補判別部は、この各領域に対して全体映像の大きさに相対的な大きさが所定の値より大きい場合に、その領域を目的候補領域から取り除くことを特徴とする請求項2に記載の話し手位置検出装置。

【請求項9】 前記テクスチャ抽出部は、各領域のテクスチャ応答を算出する最小形態論フィルタを備えた形態論解析部を有し、前記目候補判別部は、この各領域のテクスチャ応答を算出して得られたテクスチャ特性値が所定の値より小さい

場合に、その領域を目の候補領域から取り除くことを特徴とする請求項2に記載の話し手位置検出装置。

【請求項10】 前記テクスチャ抽出部は、各領域の水平エッジを検出する微分フィルタを備えた水平エッジ解析部を有し、

前記目候補判別部は、この各領域の水平エッジを検出して得られた水平エッジ特性値が所定の値より小さい場合に、その領域を目の候補領域から取り除くことを特徴とする請求項2に記載の話し手位置検出装置。

$$d^2(\mathbf{x}) = (\mathbf{x} - \boldsymbol{\mu})^t \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu}) \quad \cdots (1)$$

前記式(1)中、 \mathbf{x} は入力色差信号 C_r 、 C_b から構成されたベクトル値であり、 $\boldsymbol{\mu}$ は予め学習された皮膚色の平均ベクトルであり、 $\boldsymbol{\Sigma}$ は学習値の分散行列を示す。

【請求項13】 前記エッジ特徴解析部は、前記探索テンプレートで推定される口と鼻の各々の位置に対応するように入力されたRGB映像における第1水平エッジ大きさと、目、鼻及び口の位置以外に前記探索テンプレートとマッチングする領域に対応する入力されたRGB映像における第2水平エッジ大きさとを検出し、前記第1水平エッジ大きさと前記第2水平エッジ大きさととの比であるエッジ成分比を算出することを特徴とする請求項3に記載の話し手位置検出装置。

【請求項14】 前記エッジ特徴解析部は、目の部分を示す円の大きさによって正規化された目の水平エッジ大きさを検出し、

前記最適テンプレート決定部は、前記正規化確率距離、前記エッジ成分比及び前記正規化された目の水平エッジ大きさの各々に所定の重みを与えて得られた加重値を足し合した値が最小となるテンプレートを最適のテンプレートと決定することを特徴とする請求項13に記載の話し手位置検出装置。

【請求項15】 前記最適テンプレート決定部は、複数の探索テンプレートが重畳されて形成された領域と他の探索テンプレートが重畳されて形成された領域とが独立して位置する場合に、この独立して位置する領域ごとに別々の最適テンプレートを決定することを特徴とする請求項3に記載の話し手位置検出装置。

【請求項16】 前記最小外廓四角形の大きさ及び動き速度を算出することによって、前記形状メモリ更新部で増加または減少させるべき値の大きさを制御する速度及び形状解析部をさらに備えることを特徴とする請求項4に記載の話し手位置検出装置。

【請求項17】 下記段階(a)から段階(c)を含むことを特徴とする話し手位置検出方法。

(a) 入力されたRGB映像で、相対的に強い灰色の特徴を示すピクセルを検出し、この検出されたピクセルが形成する領域の中から部分凝集特性を示してテクスチャ特徴を有する領域を目の候補領域と決定する段階。

(b) 前記目の候補領域の中から抽出した2個の領域の

【請求項11】 前記モデルテンプレートは、両目を示す二つの円を含み、かつ鼻の部分と口の部分を挟んで四角形の底辺が位置する顔領域の四角形として構成されることを特徴とする請求項3に記載の話し手位置検出装置。

【請求項12】 前記確率距離 d は、下記式(1)によって求められることを特徴とする請求項3に記載の話し手位置検出装置。

【数1】

位置にマッチングさせて探索テンプレートを形成し、前記RGB映像で該探索テンプレートの領域内におけるピクセルの色度値に対する確率距離の和を正規化した値と、探索テンプレートによって推定される目、口及び鼻の位置における水平エッジの大きさをを用いて形成された探索テンプレート中、最適の探索テンプレートを決定する段階。

(c) 前記最適の探索テンプレートを利用して最小外廓四角形を構成し、形状メモリに記憶された各ピクセルに対してカウント値の中から前記最小外廓四角形の領域に対応するカウント値については増加させ、前記最小外廓四角形以外の領域は対応するカウント値については減少させることによって、前記形状メモリから所定の値以上のカウント値が位置する領域を話し手位置領域として出力する段階。

【請求項18】 前記段階(a)は、下記段階(a1)から段階(a5)を備えることを特徴とする請求項17に記載の話し手位置検出方法。

(a1) 入力されたRGB映像を解析して相対的に強い灰色の特徴を示すピクセルを抽出する段階。

(a2) 抽出されたピクセル中、隣接したピクセルと相互に連結されたピクセル間同士を結合して領域を形成する段階。

(a3) 形成された前記領域の各々に対して部分凝集特性を検出する段階。

(a4) 形成された前記領域の各々に対してテクスチャ特徴を検出する段階。

(a5) 形成された前記領域の中で前記部分凝集特性及び前記テクスチャ特徴が各々所定の値より大きい領域を目の候補領域と判別する段階。

【請求項19】 前記段階(b)は、下記段階(b1)から段階(b4)を備えることを特徴とする請求項17に記載の話し手位置検出装置。

(b1) 予め備わったモデルテンプレートを目の候補領域の中から抽出した2個の領域の位置に対応させて相似変換することにより入力されたRGB映像上に探索テンプレートを形成する段階。

(b2) RGB映像で探索テンプレートの領域内におけるピクセルの色度値に対する確率距離の和を前記探索テン

プレートの大きさに対して正規化した正規化確率距離を算出する段階。

(b3) 前記探索テンプレートで推定される目、鼻及び口の位置で入力されたRGB映像の水平エッジ特徴値を検出する段階。

(b4) 前記顔テンプレート形成部により形成された複数の探索テンプレート中、前記正規化確率距離及び前記水平エッジ特徴値に所定の重みを与えて得られた加重値により最適のテンプレートを決定する段階。

【請求項20】 前記(c)段階は、下記段階(c1)から段階(c3)を備えることを特徴とする請求項17に記載の話し手位置検出方法。

(c1) 前記最適の探索テンプレート内で顔映像が含まれる最小外廓四角形を構成する段階。

(c2) 形状メモリに記憶された各ピクセルに対してカウント値の中から前記最小外廓四角形の領域に対応するカウント値については増加させ、前記最小外廓四角形以外の領域に対応するカウント値については減少させる段階。

(c3) 前記形状メモリで所定の値以上のカウント値が位置する領域を話し手位置領域として出力する段階。

【請求項21】 前記(a1)段階は、前記RGB映像の各ピクセルに対してカラーを示す色要素の最大値と最小値との差が所定の値より小さく、かつ最大値が他の所定の値より小さく分布する場合に、相対的に強い灰色性向を示すピクセルとして抽出することを特徴とする請求項18に記載の話し手位置検出方法。

【請求項22】 前記(a3)段階は、各領域の環状性値を算出し、

前記(a5)段階は、領域の環状性値が所定の値より少ない場合に、その領域を目の候補領域から取り除くことを特徴とする請求項18に記載の話し手位置検出方法。

【請求項23】 前記(a3)段階は、各領域の細長比

$$d^2(x) = (x - \mu)^t \Sigma^{-1} (x - \mu) \quad \dots (1)$$

前記式(1)中、 x は入力色差信号 C_r 、 C_b から構成されたベクトル値であり、 μ は予め学習された皮膚色の平均ベクトルであり、 Σ は学習値の分散行列を示す。

【請求項29】 前記(b3)段階は、前記探索テンプレートで推定された口の位置と鼻の位置に対応して入力されたRGB映像における第1水平エッジ大きさと、目、鼻及び口の位置以外に前記探索テンプレートとマッチングする領域に対応して入力されたRGB映像における第2水平エッジ大きさとを検出し、第1水平エッジ大きさと第2水平エッジ大きさとを正規化して得られるエッジ成分比を算出することを特徴とする請求項19に記載の話し手位置検出方法。

【請求項30】 前記(b3)段階は、目部分を示す円の大きさにより正規化された目の水平エッジ大きさを検出する段階をさらに備えて、

を算出し、

前記(a5)段階は、領域の細長比が所定の値より大きい場合、または他の所定の値より小さい場合に、その領域を目の候補領域から取り除くことを特徴とする請求項18に記載の話し手位置検出方法。

【請求項24】 前記(a3)段階は、各領域に対して全体映像の大きさに相対的な領域の大きさを算出し、前記(a5)段階は、前記各領域に対して全体映像の大きさに相対的な大きさが所定の値より大きい場合に、その領域を目の候補領域から取り除くことを特徴とする請求項18に記載の話し手位置検出方法。

【請求項25】 前記(a4)段階は、各領域のテキストチャ应答を算出し、

前記(a5)段階は、この各領域のテキストチャ应答を算出して得られたテキストチャ特性値が所定の値より小さい場合に、その領域を目の候補領域から取り除くことを特徴とする請求項18に記載の話し手位置検出方法。

【請求項26】 前記(a4)段階は、各領域の水平エッジを検出し、

前記(a5)段階は、各領域の水平エッジを検出して得られた水平エッジ特性値が所定の値より小さい場合に、その領域を目の候補領域から取り除くことを特徴とする請求項18に記載の話し手位置検出方法。

【請求項27】 前記モデルテンプレートは、両目を示す二つの円を含み、かつ鼻の部分と口の部分とを挟んで四角形の底辺が位置する顔領域の四角形で構成されることを特徴とする請求項19に記載の話し手位置検出方法。

【請求項28】 前記確率距離 d は、下記式(1)によって求められることを特徴とする請求項19に記載の話し手位置検出方法。

【数1】

前記(b4)段階は、前記正規化確率距離、前記エッジ成分比及び前記正規化された目の水平エッジ大きさの各々に所定の重みを与えて得られた加重値を足し合せた値が最小となるテンプレートを最適のテンプレートと決定することを特徴とする請求項29に記載の話し手位置検出方法。

【請求項31】 前記(b4)段階は、複数の探索テンプレートが重畳されて形成された領域が他の探索テンプレートが重畳されて形成された領域と独立して位置する場合に、独立した領域ごとに別々の最適テンプレートを決定することを特徴とする請求項19に記載の話し手位置検出方法。

【請求項32】 前記(c1)段階の後に、前記最小外廓四角形の大きさ及び動き速度を算出し、前記形状メモリで増加、または減少させるべき値の大きさを決定する

段階をさらに備えることを特徴とする請求項20に記載の話し手位置検出方法。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】本発明は、映像信号処理に係り、特に、カメラセンサ等の入力手段を通して入力された映像または、ビデオ映像で示される人の顔の特徴を表示する特徴値を解析して抽出し、映像内で人の顔の位置を検出する装置及びその方法に関する。

【0002】

【従来の技術】最近、人工知能分野に関する研究は、人が有する認識能力をコンピュータに移植してコンピュータや機械に知的能力を付与することができる部分に関心と研究とが集中している。その中で、特に、人の視覚システムを利用した顔の認識技術は、映像処理、パターン認識及び顔つきの研究などのコンピュータビジョン及び映像処理の関連分野の全般に渡り非常に活発で広範囲に研究されている。

【0003】所定の入力手段で入力された映像を利用して顔の存在の有無、及び顔の領域を検出する技術は顔の認識の問題のみならず、顔つきの特徴の研究、乗り物の運転手の眠気感知、所定の場所の出入管理、映像インデクシング等の種々の応用分野で非常に重要であると認識されている。人は、各種の動的な環境下でも容易に顔の領域を探し出すことができるが、コンピュータを利用してこのような仕事を実行することは、たとえ顔の背景が比較的シンプルな構成である映像であっても、単純な問題とはならない。

【0004】既に提案されている顔の領域を検出する方法の中で代表的な方法としては、米国特許番号US5680481に記載されている神経網(neural network:ニューラルネットワーク)を利用する方法、米国特許番号US5710833に記載されている明るさの主成分の分析等による顔の明るさの統計的特徴を利用する方法、及びT. Poggioが発表した学術論文[IEEE TPAMI 20、1998]に記載されている整合方法などが挙げられる。

【0005】ところで、所定の方法によって特定の要素を抽出されて構成された顔の候補を有する映像を、顔の認識システムの入力に用いるためには、抽出された顔の候補領域における顔の構成要素や、顔の特徴に基づき顔の正確な位置を検出する過程が必要となる。

【0006】すなわち、所定の入力手段で入力された入力映像を既設のモデルと比較するためには、この入力映像から抽出された顔映像とモデルテンプレートの顔映像との間の大きさの差、回転方向の差及び位置の差などを補償するための位置抽出及び大きさ正規化過程が、認識及びマッチング性能を向上させるために必須な先決過程として要請されている。

【0007】大部分の顔認識システムでは、このような

整列及び正規化過程に必要な基準になる顔構成要素として、目の領域、またはひとみの中心領域が用いられている。その理由は、顔の構成要素中、顔映像の大きさ、表情、姿勢、照明などに変化が生じても相対的にその特徴の変化が小さい部分、すなわち相対的な変化が小さい形態を有する部分が目の領域であるためである。

【0008】このような目の領域、またはひとみ中心の位置を映像で探すための多くの研究が進められてきた。既存の顔認識システムに適用された方法が主にひとみの検出方法にも適用されてきた。ひとみ検出方法中、代表的なものとしては、様々な大きさの目のテンプレートを作って入力映像のガウスピラミッド(Gaussian pyramid)映像を構成し、入力映像内のあらゆる位置で正規化相関係数法を利用してひとみを探索する方法が一般に適用されてきた(例えば、T. Poggioが発表した学術論文[IEEE TPAMI 20、1998]に記載されている。)

【0009】米国特許番号US5680481とMoghaddamが発表した学術論文[IEEE TPAMI 19、1997]とに各々記載されている方法は、テンプレートの大きさにより目、鼻、口の領域に対する固有行列を作って、テンプレート映像内の全領域で入力映像との比較を通して関心度の比較的高い領域の特徴を探索するものである。

【0010】このような2種の方法で生じる共通的な問題としては、映像で目や鼻などの特徴に対する大きさや方向、位置等について未知であるために、これらを映像の全領域に対して大きさや方向により分類されたいくつかのモデルテンプレートを用いて探索しなければならないことが挙げられる。このことは、膨大な計算量を発生させ、また位置決定のための臨界値決定、過度の誤応答処理などの問題によって実際にシステムに適用することを困難にしている。

【0011】米国特許番号US5832115では、同心円状の二つの楕円で構成された顔の境界を検出するためのテンプレートを備え、この二つの楕円間の領域で顔を包むエッジの大きさの評価を通して顔の位置を検出するものである。しかしながら、この方法によっても、楕円のテンプレートの大きさ及び方向を決定しなければならない、全映像領域から顔の位置を探索しなければならないという問題がなお存在している。

【0012】このような顔の位置の検出に関する問題を解決するために行われている研究の中で、最近では、多くの研究がカラー映像を利用する部分に集中してきた。大部分のカラー映像では、顔色や皮膚色のカラー値が一般的な統計値の範囲に比較的近いという事実に基づいて、皮膚色を検出することによって顔の候補領域を抽出するといった研究が主流をなしている(例えば、J. Rehgが発表した学術論文:COMPAQ TR CR L9811、1998に記載されている。)

【0013】このような研究は、最近、カラーインデクシングや顔追跡及び抽出等に比較的成功的に適用されてきた。しかしながら、色による顔の位置の抽出では、顔色が映像を獲得したカメラの違いや照明色、物体表面及び状態などの映像獲得条件から多くの影響を受ける。すなわち、他のカメラで得られた映像は、たとえ同一環境条件下の同一人物に対しても異なるカラー値を有する顔色を与え易い。

【0014】とくに、照明の違いによって顔や皮膚のカラー値は、かなり大幅に変化するため、映像獲得条件が不明である場合には、顔の色領域のみを決定するために皮膚色のカラー値に対して一定の範囲を決定することが甚だ困難となる。また、背景の領域を含みながら幅広く抽出された類似した皮膚色を有する映像に対して顔の領域のみを決定する過程は容易ではなく、多くの後処理過程を必要とする。

【0015】

【発明が解決しようとする課題】本発明は前記した問題点を解決するために創作されたものであり、本発明の目的は、映像で話し手の目と顔の位置とをより正確に、しかも迅速に検出することができて、さらに映像のノイズにも強い耐性を備えた話し手位置検出装置及びその方法を提供することにある。

【0016】

【課題を解決するための手段】前記した目的を達成するために、本発明に係る話し手位置検出装置の一態様は、入力されたRGB映像で、相対的に強い灰色の特徴を示すピクセルを検出し、この検出された相対的に強い灰色の特徴を示すピクセルが形成する領域の中から部分凝集特性を示してテクスチャ特徴を有する領域を目の候補領域として決定する目位置検出手段と、前記目の候補領域の中から抽出した2個の領域の位置にマッチングさせて探索テンプレートを形成し、RGB映像で前記探索テンプレートの領域内におけるピクセルの色度値に対する確率距離の和を正規化した値と、前記探索テンプレートによって推定される目、口及び鼻の位置における水平エッジの大きさをを用いて形成された探索テンプレート中、最適の探索テンプレートを決定することによって顔の位置を決定する顔位置決定手段と、前記最適の探索テンプレートを利用して最小外廊四角形を構成し、形状メモリに記憶された各ピクセルに対してカウント値の中から前記最小外廊四角形の領域に対応するカウント値については増加させ、前記最小外廊四角形以外の領域は対応するカウント値については減少させることによって、前記形状メモリで所定の値以上のカウント値が位置した領域を話し手位置領域として出力する抽出位置安定化手段とを含むように構成する。

【0017】前記の他の目的を達成するために、本発明に係る話し手位置検出方法の一態様は、(a)入力されたRGB映像で、強い灰色の特徴を示すピクセルを検出

して、検出されたピクセルが形成する領域の中から部分凝集特性を示してテクスチャ特徴がある領域を目の候補領域と決定する段階と、(b)前記目の候補領域の中から抽出した2個の領域の位置にマッチングさせて探索テンプレートを形成して、RGB映像で探索テンプレートの領域内におけるピクセルの色度値に対する確率距離の和を正規化した値及び、探索テンプレートにより推定される目、口及び鼻の位置における水平エッジの大きさをを用いて形成された探索テンプレート中、最適の探索テンプレートを決定する段階と、(c)前記最適の探索テンプレートを利用して最小外廊四角形を構成して、形状メモリに記憶された各ピクセルに対してカウント値の中から前記最小外廊四角形の領域に対応するカウント値については増加させ、前記最小外廊四角形以外の領域は対応するカウント値については減少させながら、前記形状メモリから所定の値以上のカウント値が位置する領域を、話し手の位置領域として出力する段階とを含むように構成する。

【0018】以下、本発明に適用される基本的な概念及び作用について説明する。本発明にあっては、まず、多様なカラー映像から獲得された数多くの顔の分析を通して共通に示される顔の代表的な特徴である目の位置を抽出するものである。顔において、通常、目は幾何学的に凹状を有するために、目の明るさは黒い瞳孔を有しない西洋人の場合でも映像でそのカラー特性は相対的に強い灰色性向を示す。

【0019】すなわち、入力されたカラーの3種の主成分(RGB)の値は目の位置では相互に似た大きさを有し、かつ明るさが極めて低いという特徴を示す。このようなカラー特性が目の代表的なカラー特徴となる。

【0020】また、目の明るさ分布を見れば、ひとみと瞳孔、そして顔色などの明るさの差が稠密に集中して分布するために、ひとみの輪郭が示される大部分の映像で、テクスチャが存在する特徴を示すようになる。

【0021】また、目の位置は、目の周囲の顔色によって囲まれているために、このようなカラー及びテクスチャの2種の特徴が集中することによって部分的に生じる特徴を示すようになる。一方、髪の毛の場合には、境界部で部分的にテクスチャと相対的に強い灰色との特徴を有しているが、このような特徴が比較的幅広く、または比較的長く示されるという特徴がある。したがって、髪の毛の場合には通常、このようなカラー及びテクスチャの2種の特徴が部分的に発生する特徴を満足に示さないものとなる。

【0022】本発明にあっては、このようなひとみと瞳孔、そして顔色の3種の重要な特徴を目の位置の初期検出のための情報として用いるものである。すなわち、目で生じる3種の特徴の結合を通して目の位置を検出し、引き続き行われる多くの後処理過程の結合と、顔認識過程を通して正確な目の位置を抽出し、このようにし

て得られた情報を利用して顔の位置を抽出する。また、このようにして抽出された顔の位置情報を利用することによって、ビデオホン等の画像伝送で特に関心度の比較的高い領域への応用に適用することが可能である。

【0023】

【発明の実施の形態】以下、添付された図面を参照して本発明を詳細に説明する。図1は、本発明に係る一例の話し手位置抽出による画質改善装置の全体構成を示したブロック構成図である。図1によると、本発明に係る一例の話し手位置抽出による画質改善装置は、入力映像で目の位置を決定する目位置検出手段10、検出された目の位置の候補点を利用して顔のテンプレートを構成し、目の位置と顔の位置とを判別するために前記顔のテンプレートと映像データとをマッチングさせて顔の位置を決定する顔位置決定手段20、及び抽出された目の位置と顔の位置とが映像で急激に変化することを抑えるための抽出位置安定化手段30を含んで構成される。

【0024】図2は、図1に示す本発明に係る一例の目位置検出手段10をより詳細に示したブロック構成図である。図2によると、本発明に係る一例の目位置検出手段10は、色変換部21と、強灰色抽出部22と、中間値フィルタリング部23と、領域形成部24と、領域形成解析部25と、テキストチャ抽出部26と、目候補判別部27とを備える。

【0025】色変換部21は、入力された映像のビデオ信号YUVを3色信号RGBに変換する。強灰色抽出部22は、映像の3色信号を解析して相対的に強い灰色の特徴を示すピクセルを抽出する。一例の目位置検出手段10では、強灰色抽出部22が、任意の一つのピクセルに対してカラーを示す色要素の最大値(maxc)と最小値(minc)との差が所定の値t1より小さく、最大値(maxc)が他の所定の値t2より小さく分布すれば強い灰色性向を示すという特徴を利用するものである。

【0026】ここで、色要素の値が0～255の階調で表現されるとき、t1は55～65の値の中から決定され、t2は90～110の値の中から決定されることが望ましい。しかし、本発明の範囲は、このような実施の形態のみに限定されるものではなく、本発明の技術的思想に基づく限りにおいて、種々の変形が可能である。すなわち、相対的に強い灰色のピクセルを抽出することができる従来公知のあらゆる方法が含まれるものと解釈されなければならない。

【0027】中間値(Median)フィルタリング部23は、抽出されたピクセルを中間値フィルタリングしてスポットノイズ(spot noise)などを取り除く。

【0028】領域形成部24は、抽出されたピクセルに対して連結されたピクセルと一緒にまとめて領域を作り、各々の領域に該当ラベルを与える。

【0029】また、図2によると、本発明に係る一例の目位置検出手段10では、領域形状解析部25a、環状性(Circularity)解析部25a、細長比解析部25b及び領域大きさ解析部25cを備える。環状性解析部25aは各々のラベル化された領域に対してその形状を解析して円形状に類似しているか否かを判別し、細長比解析部25bは各々のラベル化された領域の細長比を算出し、領域大きさ解析部25cは各々のラベル化された領域の相対的な大きさを算出して各領域の部分凝集特性を検査する。

【0030】さらに、図2によると、本発明に係る一例の目位置検出手段10の一実施の形態でテキストチャ抽出部26は、形態論解析部26a及び水平エッジ解析部26bを備える。形態論解析部26aは各領域で形態論フィルタを用いてテキストチャ応答を算出してテキストチャ特徴を検査し、水平エッジ解析部26bは水平エッジフィルタを用いて水平エッジを抽出する。該形態論フィルタとしては明るさの最小変化抽出フィルタ[M. Kunt, IEEE TCSVT, 1998]を用いることができ、また水平エッジフィルタとしては一般的な微分フィルタ(Sobel operator)を用いることができる。

【0031】目候補判別部27は、ラベル化された領域の中から部分凝集特性及びテキストチャ特徴が所定の値より大きい領域を目の候補領域として判別する。

【0032】図3は、本発明に係る一例の顔位置決定手段をより詳細に示したブロック構成図である。図3によると、本発明に係る一例の顔位置決定手段20は、顔テンプレート形成部31、確率距離演算部32、エッジ特徴解析部33及び最適テンプレート透3透34を備える。

【0033】顔テンプレート形成部31は、予め備えられたモデルテンプレートを目の候補領域の中から抽出した2個の領域の位置に対応させて相似変換することにより入力されたRGB映像上に探索テンプレートを形成する。モデルテンプレートは、両目を示す二つの円を含んで、鼻の部分と口の部分との間に四角形の底辺が位置する顔の領域の四角形で構成されることが望ましい。

【0034】確率距離演算部32は、RGB映像で探索テンプレートの領域内におけるピクセルの色差信号 C_r 、 C_g 値と予め学習された統計値とを利用して顔領域の皮膚色確率距離の和を求め、その和を探索テンプレートの大きさに対して正規化した正規化確率距離を算出する。

【0035】エッジ特徴解析部33は、探索テンプレートで推定される目、鼻及び口の位置で入力されたRGB映像の水平エッジ特徴値を検出する。エッジ特徴解析部33は、探索テンプレートで推定される口と鼻との位置に対応する入力されたRGB映像における第1水平エッジ大きさ及び目、鼻及び口の位置以外に前記探索テン

レートとマッチングさせる領域に対応する入力されたRGB映像における第2水平エッジ大きさを検出して、第1水平エッジ大きさと第2水平エッジ大きさととの比であるエッジ成分比を算出する。また、エッジ特徴解析部33は、目の部分を示す円の大きさにより正規化された目の水平エッジ大きさを検出できる。

【0036】最適テンプレート決定部34は、前記正規化確率距離、前記エッジ成分比及び前記正規化された目の水平エッジ大きさの各々に所定の重みを与えて得られた加重値を足し合せた値が最小のテンプレートを最適のテンプレートとして決定する。一方、最適テンプレート決定部34は、複数の探索テンプレートが重畳されて形成された領域が他の探索テンプレートが重畳されて形成された領域と独立して位置する場合には、独立した領域ごとに別々の最適テンプレートを決定する。これは一つの映像内に複数の顔が含まれる場合を示すためである。

【0037】図4は、本発明に係る一例の抽出位置安定化手段をより詳細に示したブロック構成図である。図4によると、本発明に係る一例の抽出位置安定化手段30は、形状メモリ43、最小外廓四角形構成部41、形状メモリ更新部42及び追跡位置抽出部44を備える。また他の例として、速度及び形状解析部45をさらに備えてもよい。

【0038】形状メモリ43は、入力されたRGB映像の大きさ（映像の横長さ×縦長さ）に対応したピクセル数のカウント値を記憶する。最小外廓四角形構成部41は、最適の探索テンプレート内で顔映像が含まれる最小外廓四角形（MBR: minimum boundary rectangle）を構成する。

【0039】探索テンプレートは、両目の相対的な位置関係によっては映像に対して四角形の方向が回転する場合もあるが、最小外廓四角形は、最適探索テンプレートによって決定された顔の外廓は含み、かつ顔が回転するか否かに関係なく映像と同一な方向に形成された四角形である。

【0040】形状メモリ更新部42は、形状メモリ43に記憶された各ピクセルに対してカウント値の中から最小外廓四角形の領域に対応するカウント値については増加させ、最小外廓四角形以外の領域に対応するカウント値については減少させる。

【0041】追跡位置抽出部44は、形状メモリ43から所定値以上のカウント値が位置した領域を話し手位置領域として出力する。一方、速度及び形状解析部45は、最小外廓四角形の大きさ及び動き速度を算出して、形状メモリ更新部42により増加または、減少される値の大きさを制御する。

【0042】以下、本発明の動作内容を詳細に説明する。まず、図2、図5及び図6（A）から図6（D）を参照して本発明に係る目の候補領域の決定過程を説明する。本発明では一般静止映像または、ビデオカラー映像

のビデオ信号を入力して用いる。入力された映像のビデオ信号YUVは、色変換部21により3色信号RGBに変換される。

【0043】本発明において、目の位置検出のために考慮すべき3種の重要な特性として、カラー映像で目が有する相対的に強い灰色性向、目近辺の位置の水平エッジ、またはテクスチャ、そして分布の凝集性が挙げられる。

【0044】図5は、このような3種の特徴を説明するための図面である。図5には16名に対して各2枚ずつ計32枚の正面顔映像とその正面顔映像を平均した1枚の映像とが示されている。図5の各映像に示されるように、両目の形状は凹状が丸い円形状として集中して示されることが分かる。このような3種の特徴の組合せを通して目の候補点を抽出することが本発明による目の候補領域を決定する過程の重要な要旨である。

【0045】図6（A）は、MPEGで用いられる代表的な4個の映像を示す図である。この映像は、主に頭と上体が見える上半身（Head & Shoulder）映像である。図6（A）に共通的に示されるように、目の部分は黒色に近い相対的に強い灰色の特徴を示している。このような灰色の特徴は、図5に示すように目の部分が凹状幾何学的な形状という点に由来するものである。

【0046】したがって、強灰色抽出部22は、任意の一つのピクセルに対してカラーを示す色要素の最大値と最小値との差が小さく、かつ明るさが低く分布すれば相対的に強い灰色性向を示すという特徴を利用して、映像の色信号から相対的に強い灰色性向を示すピクセルを抽出する。図6（B）は、このような灰色の特徴を示すピクセルを抽出したものを示す図である。図6（B）では各映像で灰色性向を有するピクセルは、白色のピクセルで元の映像に重畳して表示され、目の部分のみならず背景が相対的に暗い部分がすべて抽出されている。

【0047】このような抽出ピクセルに対する映像座標における空間分布を見ると、目の部分は顔の皮膚領域内部に凝集して示されているが、背景や頭の部分の灰色ピクセルは比較的大きな塊であったり、比較的幅広く広まったりして発生していることが分かる。すなわち、このような相対的に強い凝集性を示すという目の部分の特徴は、一貫的に表れるものである。このような部分凝集特性を用いて目の候補領域のみを抽出することができる。

【0048】強灰色抽出部22から出力されたエッジピクセルに対して中間値（median）フィルタリングと領域ラベリングされた後、領域形状解析部25によって各領域の大きさと形状などが計算されて部分凝集性を示さない領域は取り除かれる。

【0049】環状性測定では領域の方向や大きさに関係なく、形状が円に近い領域を探すことが必要である。したがって、本発明に係る一実施の形態では、このような

特徴を示す形状測定基準としてHaralick[Computer & Robot Vision, Addition-Wesley Pub., 1992]による下記式(2)及び式(3)を用いることが望ましい。

【0050】

【数2】

$$\mu = \sum \|(r_k - c_k)(\bar{r} - \bar{c})| \quad \dots (2)$$

【0051】

【数3】

$$\mu = \sum \|(r_k - c_k) - (\bar{r} - \bar{c})\| \mu_R \quad \dots (3)$$

【0052】前記式(2)及び前記式(3)中、二値 μ_R と σ_R とは各々ピクセル(r_k, c_k)であって $k=0, \dots, n$ の項で定義され、 k は形状内部のピクセルに対するインデックスであり、下記(4)は領域中心の座標である。

【0053】

【数4】

$$(\bar{r} - \bar{c}) \quad \dots (4)$$

【0054】前記式(2)及び前記式(3)によって算出された二値から μ_R/σ_R 値を測定すると、この値が形状の環状性を示す値となる。この μ_R/σ_R 値が所定の値より低い値を示せば、任意の形状になる場合が多いのでその領域は取り除かれる。なお、ここに示す本発明の望ましい一実施の形態では、この μ_R/σ_R 値が1.2と決定されたが、本発明の範囲はこれに限定されるものではない。

【0055】領域のMBR(Minimum Boundary Rectangle)を求めて細長比(横/縦)を算出してこの細長比に制限を置くことにより、映像の上下方向に少し長めの領域が取り除かれる。本発明に係る一実施の形態では、細長比が0.7より小さいか、または3.5より大きい領域は取り除かれる。また、領域内のピクセル数を数えて所定の大きさ以上の領域が取り除かれる。本発明に係る一実施の形態では、領域の大きさが(映像の横ピクセル数×映像の縦ピクセル数)/1600より大きければその領域は取り除かれる。すなわち、領域形状が比較的幅広く分布したり、大きさがあまりにも大きい形態であったり、あるいは比較的長く連結された形態を取り除く作業が実行される。しかし、本発明の範囲はこのような一実施の形態で用いた数値的な制限に限定されるものではない。

【0056】図6(C)は、形態論演算子を用いて検出したテクスチャ特徴を示すものである。図6(C)によると、明るさの稠密性の差により目の部分で比較的強い応答(明るさの大きさ)を有するテクスチャ特性が抽出されている。このようなテクスチャ特徴は、領域間の境界ではないエッジ部で強く示されるという特徴がある。

また、目の近辺では垂直方向に明るさの交差が強く生じるので、水平エッジ特性が一貫的に存在することが分かる。したがって、領域形状解析を通して抽出された目の領域候補中、水平エッジ特性とテクスチャ特性とを強く含んだ部分のみを選択することによって目の候補領域を最終的に決定することが可能である。

【0057】図6(D)は、このようにして抽出した最終的な目の候補領域の位置を示している。目は一つの顔に2個存在するので、目の位置が決定されると、これと比較すべき顔テンプレートの大きさ、方向、位置などを決定することができる。すなわち、抽出された目の候補位置によって顔テンプレートを重ねて、最終的に顔の領域を判別することにより目の位置が決定される。

【0058】図3、図7(A)から図7(C)、及び図8を参照して、顔の位置を決定する過程を説明する。図7(A)から図7(C)は、モデルテンプレートを用いて最適の探索テンプレートを決定する過程を説明する図である。図7(B)は、検出された目の候補領域と重複させてマッチングされた探索顔テンプレート(face template)の形状を示している。

【0059】モデルテンプレートは、大きさが変化できる四角形状と内部に目の位置を示す二つの円形状とで構成されている。それ以外に四角形の底辺に鼻と口が位置することができる領域が内部に含まれている。モデルテンプレートは、一旦比較する目の候補領域の位置が選定されると、映像上の位置と形状、方向及び大きさが決定されて目の候補領域上に重なる。

【0060】つぎに、重なったモデルテンプレートの内部に入る映像上の領域の色と幾何学的特徴とを調べることによって選択された目の候補領域が、実際に顔上の目を示す領域であるのか否かを決定する。モデルテンプレートは、4個の因子を有して探索テンプレートに相似変換する。すなわち、二点に対する4個の式と4個の未知数とがあるので変換因子を決定することが可能である。図7(C)は、最終的に認識された目の位置と検出された顔の領域とを示している。

【0061】次に列挙した事項は、目の位置によって決定された探索テンプレートに対して顔を認識する過程を示すものである。

(1) 人の顔は皮膚色を帯びており、人の皮膚色の分布は一般に一定の範囲を有する。照明や形状が変化すると物体の反射色は、比較的幅広く、または比較的大きく変化するが、大部分の映像に対して人の顔や皮膚色は一定値と分布とを有するものであることは多くの研究から公知となっている。このような皮膚色の分布を用いれば顔の候補領域を認識することが可能である。顔の色は、2次元色空間でガウス(Gaussian)分布を有すると仮定できる。したがって、例えば数千枚のMPEGビデオ映像で皮膚色を選択して統計値を算出することができる。この統計値を利用して図7(B)で重なった探索

テンプレート内部の領域が皮膚色に近いかなを示す確率距離を算出することが可能である。本発明に係る実施の形態では下記式(1)によって算出されるMahalanobis距離を確率距離として用いている。

lanobis距離を確率距離として用いている。

【0062】

【数1】

$$d^2(x) = (x - \mu)^t \Sigma^{-1} (x - \mu) \quad \dots (1)$$

【0063】前記式(1)において、 d は確率距離であり、 x は色差信号 C_r 、 C_b から構成される入力された色のベクトル値であり、 μ は学習された皮膚色の平均ベクトルであり、 Σ は学習値の分散行列を示している。テンプレートの大きさに対して正規化された内部の色度値に対する確率距離の和が小さいほど顔の領域となる可能性は大きくなる。

【0064】(2)四角形状の探索テンプレートで底辺の中央部近辺は、口または鼻の位置となり、この部分の水平エッジ成分は比較的大きく示される。これとは反対に、口の部分、鼻の部分と目の部分とを除いた探索テンプレートにおける残りの顔の領域部分は、比較的均一な明るさ分布を有してエッジ成分は特に存在しない。したがって、これら二つの領域における水平エッジ成分の比を求めて、これらの値を判別値として利用する。

【0065】(3)目の部分の水平エッジは相対的に大きい。したがって、目の部分を示す円の大きさによって正規化された目の水平エッジ大きさは、顔を判別するために用いることができる。

【0066】図7(B)で重なった多くの探索テンプレートに対して、以上の(1)から(3)の3つの項目の値を各々算出して、各項目の重要度により対応する重みを与えて得られた加重値を足し合せた値が最小の探索テンプレートを取るようになる。但し、複数の目の候補領域の対に対して各探索テンプレートが重なる場合には、最小応答を与える探索テンプレートのみを抽出する。

【0067】一方、複数の探索テンプレートが重畳されて形成された領域が、他の探索テンプレートが重畳されて形成された領域と独立して位置する場合には、複数の人が存在するものと判別して、独立した領域ごとに別々の最適テンプレートを決定する。このような過程を通して目の位置及び顔の位置を抽出することが可能である。図8は、上半身が示される典型的な連続したMPEG映像に対して抽出した目の位置と顔の位置とを例示したものである。

【0068】以下、図4、図7(C)を参照して、抽出安定化過程を説明する。図7(C)に示されるように、顔の認識を通して抽出された目と顔とのテンプレートは連続した映像で安定化されることが必要である。自然映像シーケンス(natural image sequences)は、映像を獲得するときの環境条件、映像入力装置内部の因子等多くの要因によって常に映像雑音を有している。

【0069】したがって、たとえ極めて短い時間に入力された映像上の連続した前後二つの映像のみを取り上げ

たとしても、これら二つの映像における画質は多くの場合に相異なる特性を示す。すなわち、このような映像雑音特性から映像の特徴値の算出が影響を受け、映像で算出された特徴値は時間軸に大きく依存して変わるようになる。

【0070】このような不安定要因から映像認識や物体抽出の性能は影響を受け、図7(C)に示される目と顔とのテンプレートの位置も連続映像では安定して抽出されることが難しいという傾向を示す。このような不安定要因を取り除くため、本発明は物体の境界を示すMBR位置情報を形状メモリに記憶させる技法を通してテンプレート抽出の不安定性問題を解決したものである。

【0071】通常、質量を有する物体は慣性(moment)を有しており、人や物体が映像で運動する際に、急激な運動変化が生じても微細な時間軸ではその運動変化がよく反映されない傾向がある。特に、例えば人の上半身のみが映ったような映像の場合に、人の空間上の位置はほぼ一定の位置に連続して示される傾向があり、このような場合には人が急激な運動変化を起こしても連続映像で見れば低速では徐々に変化するように見えることがある。

【0072】すなわち、連続した前後の映像フレームの間において、この映像の抽出位置に対する時間的な相互の相関性が存在するようになる。例えば、1秒当たり20～30枚の単位で獲得されたビデオ映像を見るとき、 i 番目のフレームでは左側に存在した物体が($i+1$)番目のフレームで突然に右側に存在するようなことは、通常起こらない。このような時間軸の相互の相関性を利用すれば映像の抽出位置を安定化させてMBRを連続的に抽出及び追跡することが可能である。このような過程は前記した形状メモリを用いる技法を利用して比較的簡単に解決することができる。

【0073】まず、映像フレームの大きさ程度のカウント値を記憶するための空間を有する形状メモリを確保し、各ピクセルに対応するカウント値を初期化する。次に、抽出された顔のMBRを n 個入力し、このMBR内部のピクセルの位置に対して割り当てられた形状メモリのカウント値を増やす。連続した次の映像で同様に抽出された顔のMBRを入力し、前記した過程と全く同じ過程を繰り返す。

【0074】もし、抽出されたMBRが連続映像で連続して類似した位置で抽出された場合には、形状メモリの対応位置におけるカウント値は引続き増えていく。このカウント値は予め決められた所定の値に到達するときまで継続的に増え、所定の値に到達すればその値を維持す

るようになる。

【0075】逆に、MBR以外の位置では形状メモリのカウント値が繰り返して減少され、0に到達すれば0を続けて維持するようになる。本発明に係る一実施の形態では、カウント値が0～7の間の値を有するようになる。このような過程を繰り返すようにすれば類似の位置で連続して抽出される物体のみ安定して、しかも正確な抽出を実行することができ、位置がランダムで、かつ反復して抽出される物体の場合には、自然に形状メモリ上で比較的低いカウント値を有するようになる。

【0076】したがって、形状メモリのカウント値を検査して予め定められたしきい値以上のカウント値を示す部分のみ物体が存在するものと判別することができる。本発明に係る一実施の形態では、3以上のカウント値を示す部分のみを顔が存在する位置として判別するように構成した。しかし、本発明の範囲は前記したカウント値の範囲及び話し手位置を決定するためのしきい値に限定されるものではない。

【0077】このような本発明に係る形状累積メモリの長所は、比較的簡単に物体抽出と位置安定化とを同時に達成することができれば、性能に対する実行速度が非常に速いものとなる。また、急激な位置の変化や、抽出されたMBRの大きさに対応して、カウントステップを減らしたり加えたりして適宜に調節することができるので、動く物体の速度、物体の重要性、または形状特性にも適応できる。すなわち、より遅く動く場合と物体の顔の大きさが比較的小さい場合には、カウントステップを比較的大きくすることが望ましい。

【0078】動画像電話機や携帯電話等の動映像通話において、人の顔の領域は最も重要な関心を集める領域ROI (region of interest) となる。したがって、抽出された顔の領域の情報をを用いてMPEG-1、MPEG-2、MPEG-4、H.263等のエンコードを用いて圧縮された映像を生成することにおいて、顔の領域の画質を改善することができる。

【0079】すなわち、本発明は関心度の比較的高い領域である顔の部分のみを比較的高い画質の画像で送り、残りの背景の領域や関心度の比較的低い領域部は、比較的低い画質の画像や比較的小さい情報量で送ることによって、全体的に伝送情報量を調整し、かつ関心度の比較的高い領域の顔の部分のみを比較的高い画質の画像に維持する装置に適切に適用することができる。

【0080】本発明の望ましい実施の形態による話し手位置検出手段は、コンピューターシステムで実行できるプログラムで作成可能であり、このようなプログラムを記録したコンピュータで読むことができる記録媒体から読み取って、汎用ディジタルコンピューターシステムで実行することができる。このような記録媒体としては、マグネチック記録媒体（例えば、ROM、フロッピー（登録商標）ディスク、ハードディスク等）、光学的判

読媒体（例えば、CDROM、DVD等）及び搬送波（例えば、インターネットを通じた伝送）のような媒体が含まれる。

【0081】以上、本発明に対してその望ましい実施の形態を中心に説明した。本発明が属する技術分野で通常の知識を有する者であれば、本発明が本発明の本質的な特性から外れない範囲内で変形された形態で具現できることは容易に理解され得るものである。それゆえ、ここに開示された実施の形態は、限定的な観点ではなく本発明の技術的思想の基本概念的な観点から考慮されなければならない。本発明の範囲は前述した説明のみに限定されるものではなく特許請求の範囲に示されており、それと同等な範囲内にあるあらゆる相違点は本発明に含まれるものとして解釈されなければならない。

【0082】

【発明の効果】本発明によれば、映像で話し手の目の位置と顔の位置とをより正確に、しかもより速く探し出すことができ、なおかつ映像雑音にも強くしてそれらを実現することができる。

【0083】すなわち、停止した背景及び動く背景を有する映像に同時に適用可能で、目の位置及び顔の位置を検出する過程で、映像全体の探索を回避することによって、高速処理及び並列処理が可能となり、運動の検出などの方法と組み合わせて信頼性をより高めた話し手の位置の抽出を可能にすることができる。

【0084】本発明は、携帯用映像電話端末機、画像会議、動画像チャットモジュール等に適用可能で、さらにこれらが適用可能なビデオホン、顔の映像の高画質保存が必要な監視システム、内容基盤映像探索などの応用分野に適切に適用できる。

【図面の簡単な説明】

【図1】本発明に係る一例の話し手位置抽出による画質改善装置の全体構成を示したブロック構成図である。

【図2】図1に示す本発明に係る一例の目位置検出手段をより詳細に示したブロック構成図である。

【図3】本発明に係る一例の顔位置決定手段をより詳細に示したブロック構成図である。

【図4】本発明に係る一例の抽出位置安定化手段をより詳細に示したブロック構成図である。

【図5】顔形状の明るさ分布を説明するための図面である。

【図6】図6(A)から(D)は、目の候補領域を検出する過程を説明するための図面である。

【図7】図7(A)から(C)は、顔の位置を検出する過程を説明するための図面である。

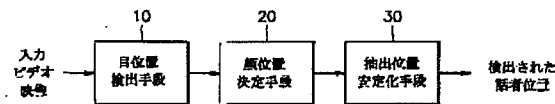
【図8】連続MPEG映像で顔の位置が検出されたことを例示的に示す図面である。

【符号の説明】

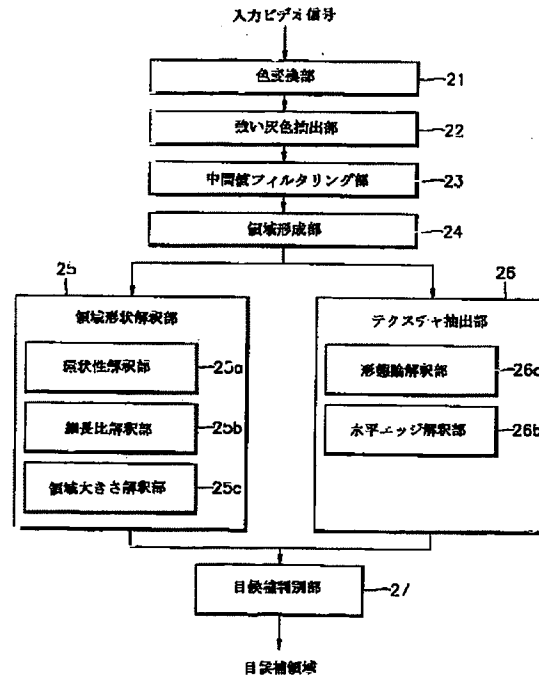
- 10 目位置検出手段
- 20 顔位置決定手段

30 抽出位置安定化手段

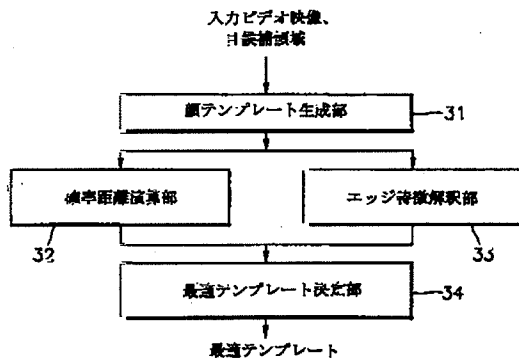
【図1】



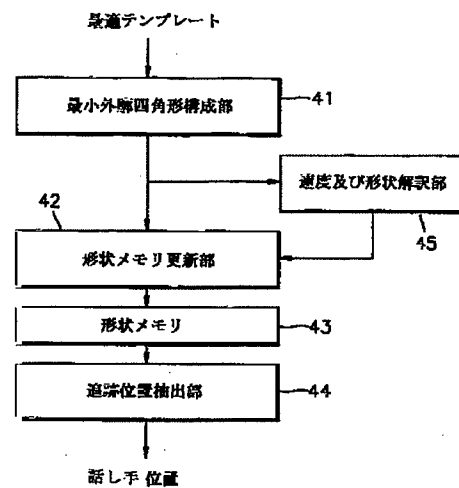
【図2】



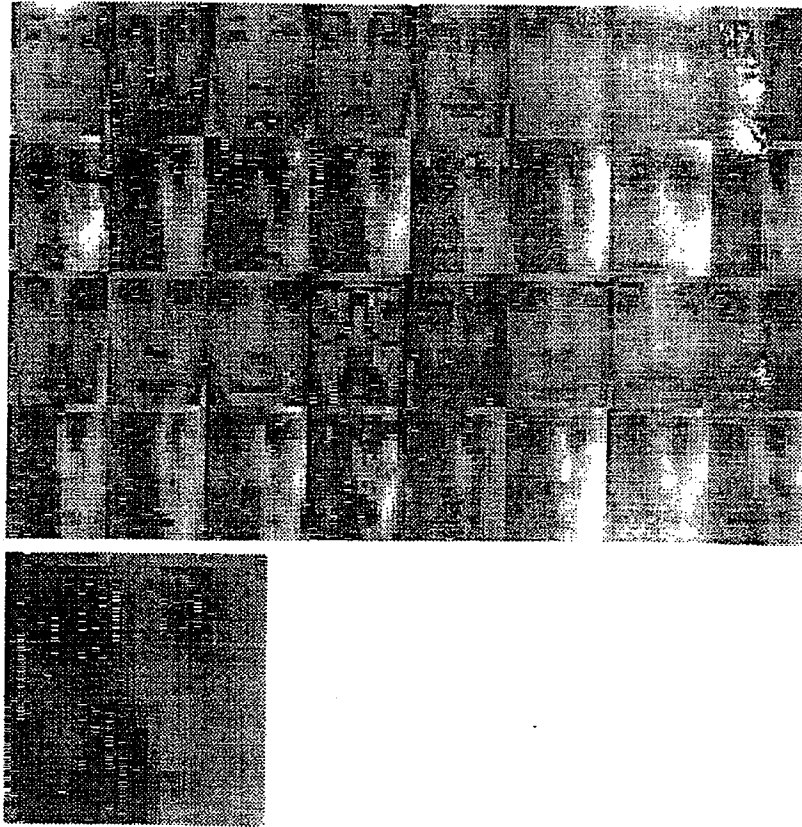
【図3】



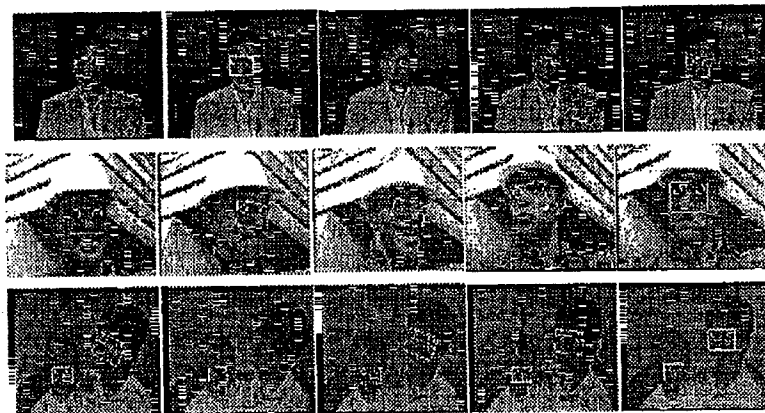
【図4】



【図5】



【図8】



【図6】

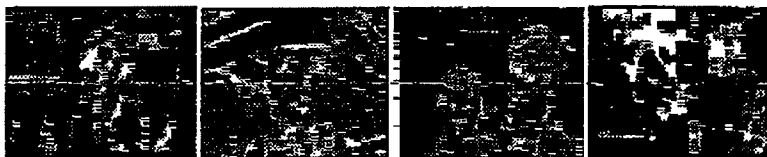
(A)



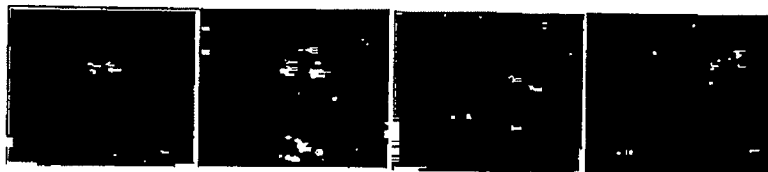
(B)



(C)



(D)

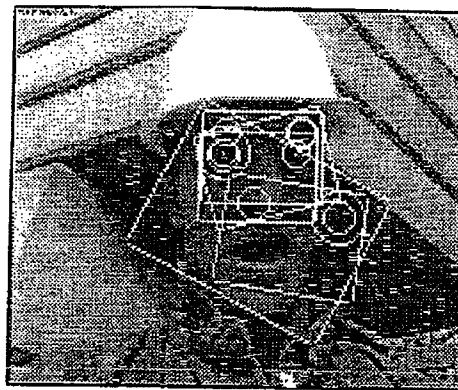


【図7】

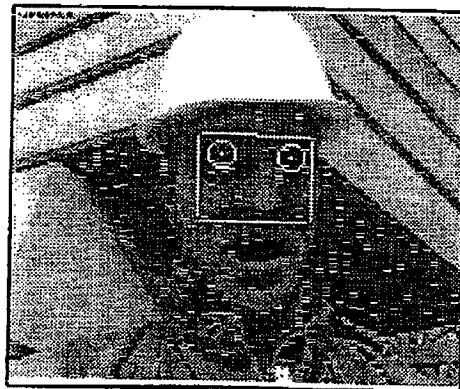
(A)



(B)



(C)



フロントページの続き

(72)発明者 金 智 淵
大韓民国 ソウル特別市 松坡区 新川洞
7番地 薔薇アパート 17棟 1101号

(72)発明者 金 昌 容
大韓民国 京畿道 儀旺市 旺谷洞 593
番地 栗谷アパート 101棟 1504号

(16) 101-175869 (P2001-175869A)

(72)発明者 徐 亮 錫

大韓民国 ソウル特別市 松波区 風納洞
219番地 美星アパート 3棟 501号

Fターム(参考) 5B057 BA02 CA01 CB01 CE06 CE18

DA08 DB06 DC09 DC16 DC25
DC32

5C066 AA11 BA20 CA21 EB03 EE02

EF11 GA01 GA02 KC08 KC09

KD06 KD07 KE04 KE07 KE24

5L096 AA02 BA18 CA04 EA45 FA06

FA15 FA18 FA32 FA33 FA69

GA55 JA09